

DATA MINING UNTUK KLASIFIKASI DENGAN ALGORITMA CART (CLASSIFICATION AND REGRESSION TREES)

Rini Astuti

Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI
Jl. Ir. H. Juanda 96 Bandung 40132

E-mail : riniastuti@likmi.ac.id

ABSTRAK

Banyaknya data hasil keluaran dari sebuah sistem informasi dapat menumpuk tidak berguna selama bertahun-tahun, dapat menjadi informasi yang lebih bermakna dengan menggunakan data mining. Data mining merupakan proses mengekstrak sekumpulan data yang dapat menambah nilai informasi berdasarkan keteraturan atau kecenderungan pola tertentu.

Salah satu teknik proses pada data mining adalah klasifikasi. Klasifikasi bertujuan untuk mengelompokkan kumpulan data berdasarkan pola atau kriteria tertentu sesuai dengan kebutuhan. Algoritma CART (Classification And Regression Trees) mengelompokkan data dengan cara membagi dua (biner) sehingga menghasilkan sebuah pohon keputusan.

Kata kunci : data mining, klasifikasi, CART.

1 PENDAHULUAN

Makin tingginya kebutuhan informasi yang lebih bermakna dari sekumpulan informasi hasil dari sebuah sistem informasi menuntut tersedianya proses untuk mengekstrak data sehingga menambah nilai informasi tersebut. Informasi yang memiliki nilai tambah dapat membantu para pengambil keputusan dalam mengambil keputusan. Salah satu cara untuk mengekstrak data dengan mengelompokkan kumpulan data dapat dilakukan dengan data mining dengan teknik klasifikasi.

Penelitian ini merupakan sebuah kajian dan studi literatur tentang data mining yang menggunakan teknik proses klasifikasi menggunakan algoritma CART (*Classification And Regression Trees*). Proses utama algoritma CART akan menghasilkan sebuah pohon keputusan yang menggambarkan hasil pengelompokan.

1.1 RUANG LINGKUP

Ruang lingkup dari data mining yang dibahas adalah tentang poses klasifikasi yang menghasilkan sebuah pohon keputusan dari data sederhana sebagai contohnya.

1.2 TUJUAN PENELITIAN

Tujuan dilaksanakannya penelitian adalah :

- a. Mengklasifikasikan suatu kumpulan data dengan teknik pohon keputusan
- b. Mengimplementasikan algoritma CART untuk suatu kumpulan data
- c. Memakai data yang jumlahnya kecil/sederhana untuk memahami tahapan algoritma CART

2 LANDASAN TEORI

Istilah data mining sudah muncul sejak tahun 1990-an dan mulai populer sejak awal tahun 2000-an. Data mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode data mining ini dapat digunakan untuk mengambil keputusan di masa depan. Data mining ini juga dikenal dengan istilah pattern recognition (Santosa, 2007) [4].

Data mining merupakan metode pengolahan data berskala besar oleh karena itu data mining ini memiliki peranan penting dalam bidang industri, keuangan, cuaca, ilmu dan teknologi. Secara umum kajian data mining membahas metode-metode seperti, clustering, klasifikasi, regresi, seleksi variable, dan market basket analisis (Santosa, 2007).

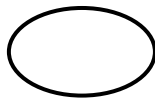
Salah satu metode untuk proses data mining adalah klasifikasi. Klasifikasi ini lebih mengacu kepada pengelompokan dengan model pohon keputusan secara biner. Salah satu algoritmanya adalah CART (*Classification And Regression Trees*).

Algoritma untuk klasifikasi lain yang sejenis misalnya algoritma C4.5 dan C5.0, algoritma *k-nearest neighbor*, algoritma ID3 (Berry dan Browne, 2006 [1]).

Menurut Yohannes [7] dan Webb (Otok, 2009: XVI-2) [3] tingkat kepercayaan yang dapat digunakan dalam pengklasifikasian data baru pada CART adalah akurasi yang dihasilkan oleh pohon klasifikasi yang murni dibentuk dari data yang mempunyai kesamaan kondisi.

CART secara rekursif memecah rekaman dalam data yang ditetapkan ke dalam himpunan bagian catatan dengan nilai yang sama untuk atribut target (Larose) [2].

Symbol Pohon keputusan yang digunakan adalah : [2]



: simpul (*node*) dasar yang merupakan keputusan yang akan bercabang dan klasifikasi belum dapat ditentukan



: simpul (*node*) terminasi kejadian yang keputusannya sudah jelas dan klasifikasi dapat ditentukan.

Langkah-langkah algoritma CART adalah sebagai berikut :

- Mempersiapkan data yang akan diklasifikasikan
- Menentukan variabel prediktor sebagai variabel sebagai dasar pengelompokan berdasarkan variabel tujuan (target)
- Menentukan calon cabang (*candidate split*) kiri dan kanan
- Mengukur *goodness* (kesesuaian) dari calon masing-masing cabang s pada simpul keputusan t yang dihitung dengan rumus :

$$\Phi(s|t) = 2 P_L P_R \sum_{j=1}^{\text{jlh kategori}} |P(j|t_L) - P(j|t_R)| \quad (2.1)$$

Keterangan :

t_L = calon cabang kiri dari node keputusan t

t_R = calon cabang kanan dari node keputusan t

P_L = jumlah data catatan pada calon cabang kiri t_L / jumlah data catatan seluruhnya

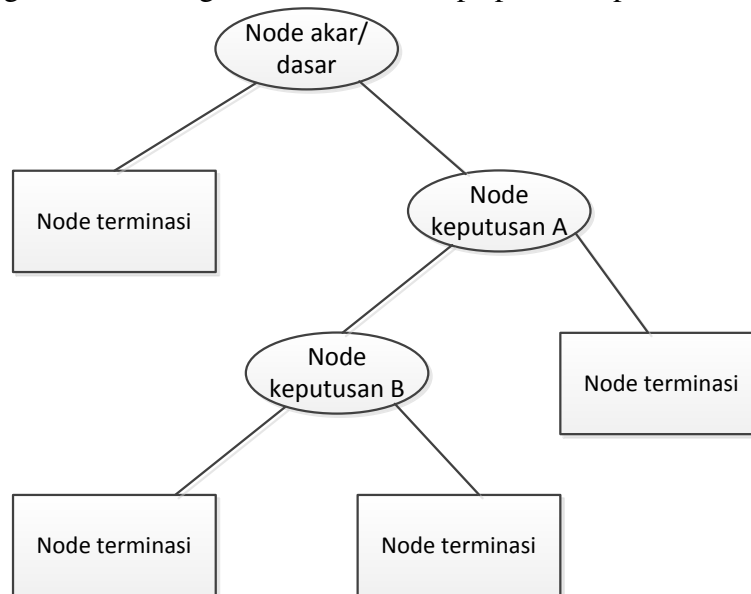
P_R = jumlah data catatan pada calon cabang kanan t_R / jumlah data catatan seluruhnya

$P(j|t_L)$ = jumlah data catatan berkategori j pada calon cabang kiri t_L / jumlah data catatan pada node keputusan t

$P(j|t_R)$ = jumlah data catatan berkategori j pada calon cabang kanan t_R / jumlah data catatan pada node keputusan t

- Tentukan calon cabang untuk node keputusan dengan memilih nilai $\Phi(s|t)$ terbesar, cabang ini tidak dihitung lagi selanjutnya.
- Gambarkan cabang node keputusan dan node kejadian terminasi.
- Ulangi langkah 4 sampai tidak terdapat lagi cabang node keputusan.

Berikut gambar hasil algoritma CART berupa pohon keputusan.



Gambar 1. Contoh Pohon keputusan

3 METODOLOGI DAN PEMBAHASAN

3.1 METODOLOGI

Metode yang akan dipakai di sini adalah studi literatur, eksplorasi data dan mengimplementasikan algoritma CART untuk suatu kumpulan data.

Berikut adalah uraian metode yang digunakan :

- a. Melakukan Studi Literatur terhadap berbagai jenis buku, jurnal, dan teori mengenai Data mining proses klasifikasi.
- b. Menggunakan kumpulan data dari sumber referensi yang ada.
- c. Mengimplementasikan Algoritma CART dengan bantuan program Microsoft Excel

3.2 PEMBAHASAN

3.2.1 KUMPULAN DATA

Kumpulan data yang digunakan adalah data tentang kendaraan yang akan dikelompokkan berdasarkan manfaatnya [5]. Berikut tabel kumpulan data untuk klasifikasi.

Tabel 1. Kumpulan data

No	Price	Maintenance	Capacity	Airbag	Profitable
1	low	low	2	no	yes
2	low	med	4	yes	no
3	low	low	4	no	yes
4	low	high	4	no	no
5	med	med	4	no	no
6	med	med	4	yes	yes
7	med	high	2	yes	no
8	med	high	5	no	yes
9	high	med	4	yes	yes
10	high	high	2	yes	no
11	high	high	5	yes	yes

3.2.2 IMPLEMENTASI ALGORITMA CART

Berdasarkan algoritma CART akan ditentukan klasifikasi kumpulan data pada Tabel 1. Dengan target profitable yes dan no. Dari tabel tersebut yang menjadi variabel prediktor adalah price, maintenance, capacity dan airbag. Selanjutnya menentukan calon cabang kiri dan kanan yang membentuk cabang biner seperti di Tabel 2.

Tabel 2. Calon cabang

Nomor Calon Cabang	Calon Cabang Kiri	Calon Cabang Kanan
1	price = low	price = (med, high)
2	price= medium	price= (low, high)
3	price=high	price= (low, med)
4	maintenance = low	maintenance = (med, high)
5	maintenance = med	maintenance = (low, high)
6	maintenance = high	maintenance = (low,med)
7	capacity =2	capacity =(4,5)
8	capacity =4	capacity =(2,5)
9	capacity =5	capacity =(2,4)
10	air bag = yes	air bag = no

Selanjutnya dari tabel 2. dihitung nilai *goodness* □□ untuk masing-masing cabang dengan rumus 2.1. ntuk mpermudah perhitungan , rumus 2.1 disisipkan sebagai berikut :

$$Q(s|t) = \sum_{j=1}^{j \text{lh kategori}} |P(j|t_L) - P(j|t_R)| \quad (3.1)$$

Maka rumus 2.1 menjadi :

$$\Phi(s|t) = 2 P_L P_R Q(s|t) \quad (3.2)$$

Misalkan untuk cabang kiri no.1 price =low dan cabang kanan price = (medium,high). Di tabel 1 terdapat 4 buah data dengan nilai price=low, sedangkan data seluruhnya adalah 11, sehingga didapat $P_L = 4/11$ atau $P_L = 0.36364$. Cabang kanan no.1 price = (medium,high) sebanyak 7 buah data dari 11 buah data seluruhnya, sehingga didapat $P_R = 7/11$ atau $P_R = 0.6364$. Selanjutnya dihitung $P(j|t_L)$ untuk profitable =Yes terdapat 2 buah untuk price =low dari 4 buah data, sehingga $P(j|t_L) = 2/4$ dan 2 buah untuk profitable=no dari 4 data, maka $P(j|t_L) = 2/4$. Dan dihitung $P(j|t_R)$ untuk profitable =Yes terdapat 4 buah dari 7 buah data untuk price =(medium,high) sehingga $P(j|t_R) = 4/7$ atau $P(j|t_R) = 0.5714$ dan 3 buah data dari 7 buah data untuk profitable=no, maka $P(j|t_R) = 3/7$ atau $P(j|t_R) = 0.4286$.

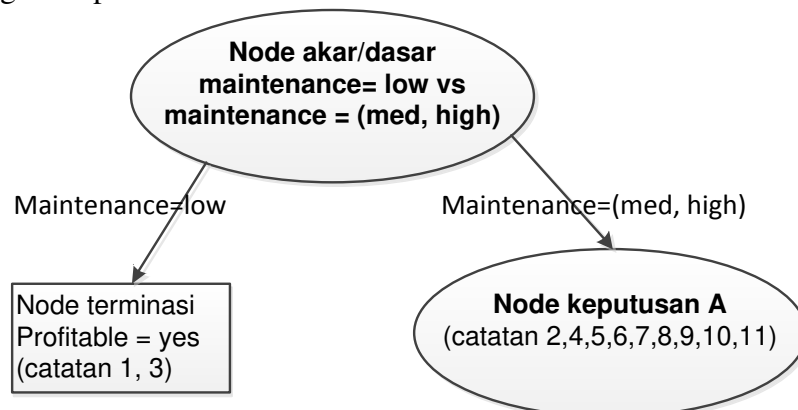
Dari persamaan 3.1 didapatkan $Q(s|t) = |0.5 - 0.5714| + |0.5 - 0.4286| = 0.1429$. Dari persamaan 3.2 didapatkan nilai □(s|t) = $2 * 0.36364 * 0.6364 * 0.1429 = 0.061$. Hasil perhitungan dilakukan untuk semua kumpulan data, selengkapnya dituangkan dalam tabel 3.

Tabel 3. Menghitung nilai *goodness* cabang

No. Calon Cabang	P_L	P_R	Profi-table	$P(j t_L)$	$P(j t_R)$	$2 * P_L * P_R$	$Q(s t)$	□(s t)
1	0.36364	0.6364	Yes	0.5000	0.5714	0.4628	0.0714	

No. Calon Cabang	P_L	P_R	Profi- table	$P(j tL)$	$P(j tR)$	$2 * P_L * P_R$	$Q(s t)$	$\square(s t)$
	4/11=	7/11=	No	0.5000	0.4286		0.0714 0.1429	0.0661
2	0.36364 4/11=	0.6364 7/11=	Yes No	0.5000 0.5000	0.5714 0.4286		0.0714 0.1429	0.0661
3	0.27273 3/11=	0.7273 8/11=	Yes No	0.6667 0.3333	0.5000 0.5000	0.3967	0.1667 0.1667 0.3333	0.1322
4	0.18182 2/11=	0.8182 9/11=	Yes No	1.0000 0.0000	0.4444 0.5556	0.2975	0.5556 0.5556 1.1111	0.3306
5	0.36364 =4/11	0.6364 =7/11	Yes No	0.5000 0.5000	0.5714 0.4286	0.4628	0.0714 0.0714 0.1429	0.0661
6	0.45455 =5/11	0.5455 =6/11	Yes No	0.4000 0.6000	0.6667 0.3333	0.4959	0.2667 0.2667 0.5333	0.2645
7	0.27273 3/11=	0.7273 8/11 =	Yes No	0.3333 0.6667	0.6250 0.3750	0.3967	0.2917 0.2917 0.5833	0.2314
8	0.54545 =6/11	0.4545 =5/11	Yes No	0.5000 0.5000	0.6000 0.4000	0.4959	0.1000 0.1000 0.2000	0.0992
9	0.18182 =2/11	0.8182 =9/11	Yes No	1.0000 0.0000	0.4444 0.5556	0.2975	0.5556 0.5556 1.1111	0.3306
10	0.54545 =6/11	0.4545 =5/11	Yes No	0.5000 0.5000	0.6000 0.4000	0.4959	0.1000 0.1000 0.2000	0.0992

Dari hasil perhitungan di Tabel 3, diperoleh nilai terbesar adalah calon cabang ke-4, sehingga pohon keputusan dari algoritma CART iterasi ke-1 terbentuk seperti pada Gambar 2. Dari tabel 1, maintenance = low menghasilkan kejelasan profitable = yes (data 1 dan 3), sehingga didapatkan node terminasi untuk maintenance=low.



Gambar 2. Pohon keputusan kumpulan data contoh untuk iterasi ke-1

Karena masih terdapat node keputusan A untuk data catatan 2,4,5,6,7,8,9,10, dan 11, maka algoritma CART diulangi untuk iterasi kedua tanpa cabang ke-4. Iterasi ke-2, data contoh tanpa no.1 dan no.3 karena sudah terdapat pada node terminasi, sehingga data yang diperhitungkan untuk menghasilkan nilai *goodness* □□cabang pada iterasi ke-2 hanya sebanyak 9 buah data yaitu untuk data catatan 2,4,5,6,7,8,9,10, dan 11. Pada iterasi ke-2, kumpulan data yang diolah seperti pada tabel 4 dan calon cabang sisa yang terbentuk seperti di tabel 5 tanpa calon cabang no.4.

Tabel 4. Kumpulan data pada iterasi ke-2

No	Price	Maintenance	Capacity	Airbag	Profitable
2	low	med	4	yes	no
4	low	high	4	no	no
5	med	med	4	no	no
6	med	med	4	yes	yes
7	med	high	2	yes	no
8	med	high	5	no	yes
9	high	med	4	yes	yes
10	high	high	2	yes	no
11	high	high	5	yes	yes

Tabel 5. Calon cabang pada iterasi ke-2

Nomor Calon Cabang	Calon Cabang Kiri	Calon Cabang Kanan
1	price = low	price = (med, high)
2	price= medium	price= (low, high)
3	price=high	price= (low, med)
4	maintenance = low	maintenance = (med, high)
5	maintenance = med	maintenance = (low, high)
6	maintenance = high	maintenance = (low,med)
7	capacity =2	capacity =(4,5)
8	capacity =4	capacity =(2,5)
9	capacity =5	capacity =(2,4)
10	air bag = yes	air bag = no

Selanjutnya menghitung nilai *goodness* □□cabang, misalkan untuk cabang kiri no.2 price = medium dan cabang kanan price = (low,high). Di tabel 4 terdapat 4 buah data dengan nilai price=medium, sedangkan data seluruhnya adalah 9, sehingga didapat $P_L = 4/9$ atau $P_L = 0.4444$. Cabang kanan no.2 price = (low,high) sebanyak 5 buah data dari 9 buah data seluruhnya, sehingga didapat $P_R = 5/9$ atau $P_R = 0.5556$. Selanjutnya dihitung $P(j|t_L)$ untuk profitable =Yes terdapat 2 buah untuk price =medium dari 4 buah data, sehingga $P(j|t_L) = 2/4$ dan 2 buah untuk profitable=no dari 4 data, maka $P(j | t_L) = 2/4$. Dan dihitng $P(j|t_R)$ untuk profitable =Yes terdapat 2 buah dari 5 buah data untuk price =(low,high) sehingga $P(j | t_R) = 2/5$ atau $P(j | t_R) = 0.4$ dan 3 buah data dari 5 buah data untuk profitable=no, maka $P(j | t_R) = 3/5$ atau $P(j | t_R) = 0.6$.

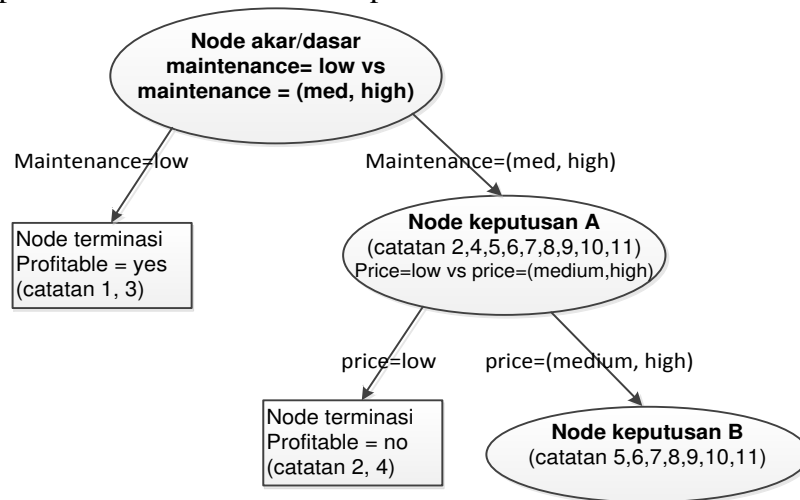
Dari persamaan 3.1 didapatkan $Q(s|t) = |0 = 0.5 - 0.4| + |0.5 - 0.6| = 0.2$. Dari persamaan 3.2 didapatkan nilai $\square(s|t) = 2 * 0.4444 * 0.5556 * 0.2 = 0.0988$. Hasil

perhitungan dilakukan untuk semua kumpulan data (9 buah), selengkapnya dituangkan dalam tabel 6.

Tabel 6. Menghitung nilai *goodness* □ □ cabang pada iterasi ke-2

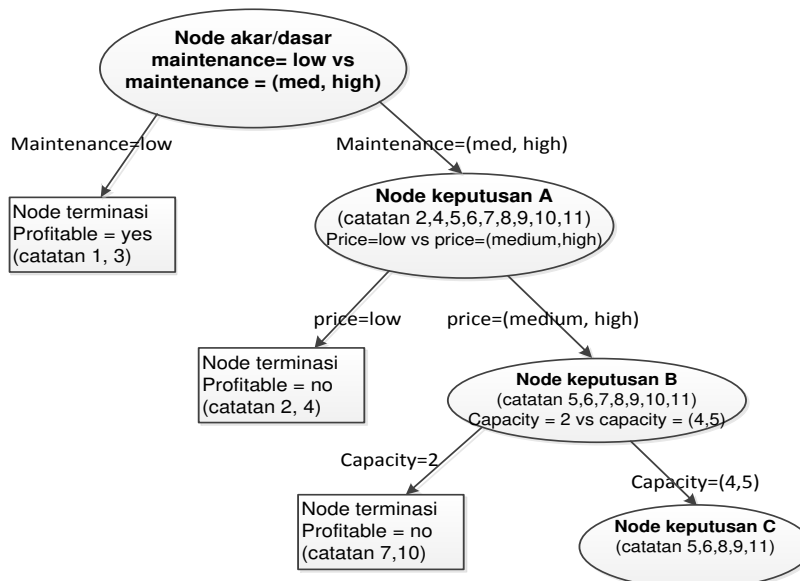
Nomor Calon Cabang	P_L	P_R	profit able	$P(j tL)$	$P(j tR)$	$2^* P_L$ $* P_R$	$Q(s t)$	$\square(s t)$
1	0.2222 =2/9	0.7778 7/9=	Yes No	0.0000 1.0000	0.5714 0.4286	0.3457	0.5714 0.5714 1.1429	0.3951
2	0.4444 4/9=	0.5556 5/9=	Yes No	0.5000 0.5000	0.4000 0.6000	0.4938	0.1000 0.1000 0.2000	0.0988
3	0.3333 3/9=	0.6667 6/9=	Yes No	0.6667 0.3333	0.3333 0.6667	0.4444	0.3333 0.3333 0.6667	0.2963
4								
5	0.4444 =4/9	0.5556 =5/9	Yes No	0.4000 0.6000	0.4000 0.6000	0.4938	0.0000 0.0000 0.0000	0.0000
6	0.5556 =5/9	0.4444 =4/9	Yes No	0.4000 0.6000	0.5000 0.5000	0.4938	0.1000 0.1000 0.2000	0.0988
7	0.2222 = 2/9	0.7778 =7/9	Yes No	0.0000 1.0000	0.5714 0.4286	0.3457	0.5714 0.5714 1.1429	0.3951
8	0.5556 =5/9	0.4444 =4/9	Yes No	0.4000 0.6000	0.4000 0.5000	0.4938	0.0000 0.1000 0.1000	0.0494
9	0.2222 =2/9	0.7778 =7/9	Yes No	1.0000 0.0000	0.2857 0.7143	0.3457	0.7143 0.7143 1.4286	0.4938
10	0.6667 =6/9	0.3333 =3/9	Yes No	0.5000 0.5000	0.6000 0.4000	0.4444	0.1000 0.1000 0.2000	0.0889

Dari hasil perhitungan di Tabel 6, diperoleh nilai terbesar adalah calon cabang ke-1, sehingga pohon keputusan dari algoritma CART iterasi ke-2 terbentuk seperti pada Gambar 3. Dari tabel 4, price = low menghasilkan kejelasan profitable = no (data 2 dan 4), sehingga didapatkan node terminasi untuk price=low.



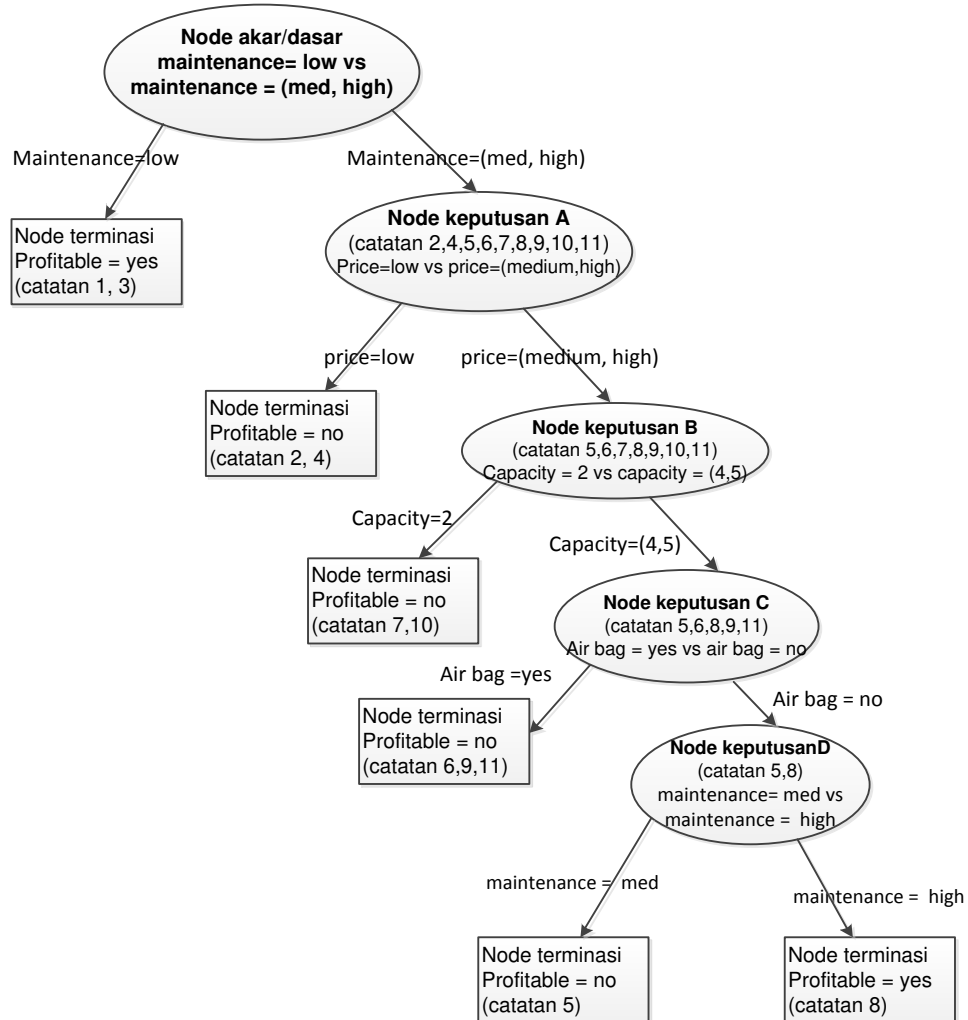
Gambar 3. Pohon keputusan kumpulan data contoh untuk iterasi ke-2

Karena masih terdapat node keputusan B untuk data catatan 5,6,7,8,9,10, dan 11, maka algoritma CART diulangi untuk iterasi ke-3 tanpa cabang ke-1. Iterasi ke-3, data contoh tanpa no.2 dan no.4 karena sudah terdapat pada node terminasi, sehingga data yang diperhitungkan untuk menghasilkan nilai *goodness* □□ cabang pada iterasi ke-3 hanya sebanyak 7 buah data yaitu untuk data catatan 5,6,7,8,9,10, dan 11. Dengan cara yang sama, pada iterasi ke-3, diperoleh nilai terbesar adalah calon cabang ke-7, sehingga pohon keputusan dari algoritma CART iterasi ke-3 terbentuk seperti pada Gambar 4. Dari tabel 4, capacity = 2 menghasilkan kejelasan profitable = no (data 7 dan 10), sehingga didapatkan node terminasi untuk capacity=2.



Gambar 4. Pohon keputusan kumpulan data contoh untuk iterasi ke-3

Karena masih terdapat node keputusan C untuk data catatan 5,6, 8,9,11, maka algoritma CART diulangi untuk iterasi ke-4 tanpa cabang ke-7. Pada iterasi ke-4, diperoleh nilai terbesar adalah calon cabang ke-10, sehingga pohon keputusan dari algoritma CART iterasi ke-4 terbentuk seperti pada Gambar 5. Dari tabel 4, air bag = yes menghasilkan kejelasan profitable = yes (data 6,9 dan 11), sehingga didapatkan node terminasi untuk air bag = yes.



Gambar 5. Pohon keputusan kumpulan data contoh untuk iterasi ke-4 selesai

Setelah iterasi ke-4, catatan data tinggal 2 yaitu no. 5 dan no. 8 yang menghasilkan kejelasan profitable = yes untuk maintenance = high (data no. 8) dan profitable = no untuk maintenance = med (data no.5). Dengan demikian algoritma CART selesai, dan terbentuklah klasifikasi dengan hasil pohon keputusan.

3.2.3 ALAT BANTU UNTUK DATA MINING

Terdapat beberapa *tools* atau alat bantu berupa program aplikasi yang dapat digunakan untuk proses data mining secara gratis (*open source*) selain aplikasi statistik SPSS. Berikut contoh alat bantu untuk data mining : [6]

- a. **Weka** : adalah perangkat lunak bebas dan *open source* berbasis Java yang dilisensikan di bawah GNU GPL dan tersedia untuk digunakan di Linux, Mac OS X dan Windows.
- b. **Rapid Miner** : tersedia dalam edisi *open source* dan komersial dan merupakan platform analitik prediktif terkemuka. Gartner, perusahaan riset dan penasihat AS, telah mengakui Rapid Miner sebagai platform analitik canggih pada 2016. Rapid Miner

membantu perusahaan melakukan analisis prediktif dalam proses bisnis mereka dengan pustaka yang kaya akan ilmu data dan algoritma *machine learning* melalui lingkungan pemrograman lengkapnya seperti Rapid Miner Studio. Selain fitur data mining standar seperti pembersihan data, pemfilteran, pengelompokan, dll, perangkat lunak ini juga dilengkapi dengan template bawaan, alur kerja berulang, lingkungan visualisasi profesional, dan integrasi tanpa batas dengan bahasa seperti Python dan R ke dalam alur kerja yang membantu dengan cepat pembuatan prototipe. Rapid Miner digunakan untuk aplikasi bisnis / komersial, penelitian dan pendidikan.

- c. **Orange** : Pengguna Python mencoba dengan data sains yang mungkin sesuai dengan aplikasi Orange. Ini adalah pustaka Python yang mendukung skrip Python dengan kompilasi yang kaya dari data mining dan algoritma *machine learning* untuk pra-pemrosesan data, klasifikasi, pemodelan, regresi, pengelompokan dan fungsi lain-lain. Orange juga dilengkapi dengan lingkungan pemrograman visual dan meja kerjanya terdiri dari alat untuk mengimpor data, dan *drag and drop widget* dan tautan untuk menghubungkan berbagai *widget* untuk menyelesaikan alur kerja.
- d. **Apache Mahout** : Mahout terutama untuk algoritma *machine learning* yang dapat membantu dalam pengelompokan, klasifikasi dan pola mining yang sering. Ini dapat digunakan dalam mode terdistribusi yang membantu integrasi yang mudah dengan Hadoop.
- e. **ELKI** : perangkat lunak *open source* yang ditulis dalam Java dan dilisensikan dengan AGPLv3. Perangkat lunak ini berfokus terutama pada analisis kluster dan deteksi outlier dengan kompilasi berbagai algoritma dari kedua domain ini. Perangkat lunak diakses melalui GUI yang menampilkan hasil setelah algoritma yang dipilih dijalankan.

Tool lainnya adalah **MOA** - Massive Online Analysis, didistribusikan di bawah GNU GPL dan dapat digunakan melalui baris perintah, GUI atau JAVA API. **KEEL** (Knowledge Extraction for Evolutionary Learning) adalah *open source* berbasis di bawah lisensi GPLv3. Kemudian **RATTLE**, kependekan dari '*R Analytical Tool To Learn Easily*', dibangun menggunakan bahasa pemrograman statistik R. Dapat dipakai di sistem operasi Linux, Mac OS and Windows.

4 KESIMPULAN

Kesimpulan untuk Data mining proses kasifikasi dengan algoritma CART menggunakan data sederhana adalah sebagai berikut :

- a. Algoritma CART dengan mudah dipahami untuk dapat digunakan sebagai proses klasifikasi sekumpulan data dengan hasil pohon keputusan.
- b. Banyaknya iterasi di algoritma CART bergantung pada sekumpulan data dan variabel prediktor yang diproses untuk klasifikasi.
- c. Untuk sekumpulan data contoh yang dipakai disini, jika maintenance = low , maka nilai manfaat/profitable adalah yes. Jika dengan airbag , price = low, capacity= 2 dan maintenance= med, maka nilai profitable adalah no.
- d. Proses data mining untuk pengklasifikasian dan lainnya terutama untuk data yang sangat besar dapat menggunakan program aplikasi seperti Rapid Miner dan lain-lain.

5 REFERENSI

- [1]. Berry, M.W dan M. Browne,2006. Lecturer Note in Data Mining Singapore, World Scientific..
- [2]. Larose, Daniel T., (2005), Discovering knowledge in data : an introduction to data mining, John Wiley & Sons, Inc

- [3]. Otok, B. W. (2009). Bagging Cart pada Klasifikasi Anak Putus Sekolah. Seminar Nasional Statistika IX. Hlm: XVI-1-XVI-9.
- [4]. Santosa, B. (2007). Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu.
- [5]. [Online] Available: <https://nptel.ac.in/courses/106106139/> Computer Science and Engineering / Introduction to Machine Learning .[Accessed 2 Desember 2018].
- [6]. [Online] Available : <https://opensourceforu.com/2017/03/top-10-open-source-data-mining-tools/> Top 10 open source data mining tools, By Shravan I.V - March 25, 2017, [Accessed 7 Januari 2019].
- [7]. Yohannes, Y. & Hoddinott, J. (1999). Classification and Regression Trees: An Introduction. International Food Policy Research Institute (IFPRI). Washington, D.C., U.S.A.